•1991 DE_____ ONS LTD

WO 9109045 A   9128

BASIC

B4 D16

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

TION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

91-208081

C12N 15/00

A1

(11) International Publication Number:	WO 91/09045
(43) International Publication Date:	27 June 1991 (27.06.91)

(21) International Application Number: PCT/US90/07333

(22) International Filing Date: 13 December 1990 (13.12.90)

(30) Priority data:
451,953   19 December 1989 (19.12.89)   US
575,524   31 August 1990 (31.08.90)   US

(71) Applicant: THE UNITED STATES OF AMERICA, represented by THE SECRETARY, UNITED STATES DEPARTMENT OF COMMERCE [US/US]; Washington, DC 20231 (US).

(72) Inventor: YANG, Stringner, Sue ; 6625 Marywood Road, Bethesda, MD 20817 (US).

(74) Agents: OLIFF, James, A. et al.; Oliff & Berridge, P.O. Box 19928, Alexandria, VA 22320 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, KR, LU (European patent), NL (European patent), SE (European patent).

Published
With international search report.

(54) Title: HEPATOCELLULAR CARCINOMA ONCOGENE

(57) Abstract

The present invention relates to an onco_____ tein specific for hepatocellular carcinomas and to a nucleotide sequence that codes for such a protein. The invention furth___ _late; to screening and diagnostic methodologies (and kits based thereon) that make use of the oncoprotein (or antibodies s___ ific for same) and the nucleotide sequence.

BEST AVAILABLE COPY

-1-

HEPATOCELLULAR CARCINOMA ONCOGENE

TECHNICAL FIELD

The present invention relates, in general, to a
protein of hepatoma cells, and, in particular, to an
5   oncoprotein that is an amplified gene expression product
of hepatoma cells. The invention further relates to a
nucleotide fragment coding for the oncoprotein, to a
recombinant molecule that includes such a fragment and to
cells transformed therewith. The invention further
10   relates to methods of detecting the presence of hepatocel-
lular carcinomas in a patient and to kits based thereon.

BACKGROUND INFORMATION

Epidemiological evidence has led to a strong
etiological implication of several DNA viruses with the
15   occurrence of certain cancers and other disorders in
humans. These include the papillomavirus in cervical
carcinoma (HPV 16) and in epidermodysplasia verruciformis
(HPV 3 and 8); the Epstein-Barr virus in Burkitt's lympho-
ma; and the hepatitis B virus (HBV) in human hepatocellu-
20   lar carcinoma (Beasley et al, In: Vyas GN, Dienstag JL,
Hoofnagle JH, eds. Viral hepatitis and liver disease.
Orlando, FL, Grune and Stratton, 1984, 209-224). These
observations, together with the correlation of retroviral
infection such a HTLV-I in Adult T-cell leukemia asserts
25   the possible r e of infectious viruses acting as trans-
ducing agents in the pathogenesis of these aforementioned
human neoplasms and disorders.

The mechanism(s) by which infectious viruses exert
their oncogenicity is believed to be mediated by DNA
30   recombination with the host cell DNA. The mammalian
genome contains certain genes, designated proto-oncogenes,
that can acquire oncogenic properties upon transduction
into the genome of acute transforming retroviruses (Bish-
op, Ann. Rev. Biochem. 1983, 52:301; Bishop, Cell 1985,
35   42:23). In certain human cancers (e.g. T24 and EJ human
bladder carcinoma) it has been well documented that the
identified transforming gene (H-ras-1 locus) relates to
the v-rasH of the Harvey murine sarcoma virus. Among the

- 2 -

proto-oncogenes and oncogenes, the ras family has been
thoroughly characterized and studied with respect to
activation and expression in human neoplasms.   When a
proto-oncogene undergoes point-mutation (e.g. c-rasH) or
5   rearrangement (e.g. n-myc), such changes can lead to a
loss of cell regulation in differentiation and growth, and
eventually oncogenesis.

Recently, a transforming DNA sequence from a human
(Mahlavu) hepatocellular carcinoma, hhc^M, has been identi-
10   fied and molecularly cloned as part of a large fragment
(Yang et al, J. Gen. Virol. 1982, 63:25; Yang et al,
Environmental Health Perspectives 1985, 62:231). A number
of hhc^M related DNA clones from several other human hepa-
tocellular carcinomas have been isolated that exhibited
15   nil to moderate cell transforming activity on NIH/3T3
cells. Two have been partially characterized and they are
a moderately cell-transforming gene from Mahlavu hepato-
cellular carcinoma (hhc^M) and a putative cellular homo-
logue (c-hhc) isolated from normal human liver DNA, which
20   has no cell-transforming activity. The biological activi-
ties of two molecular clones of hhc^M and a Korean hhc^{KJ} and
c-hhc have been characterized and compared (Yang et al,
Leukemia 1988, .2 Supplement):102S). Amplification of
the hhc^M seqr .e in the various genomic DNAs of hepatomas
25   from 2 Chinese, one African and 17 Korean sources, was
observed and compared with the distribution of integrated
HBV DNA sequences in the same hepatomas in order to
provide some insight into the possible role of hhc^M.

The present invention relates to an oncoprotein
30   specific for hepatocellular carcinomas and to a nucleotide
sequence that codes for such a protein. The invention
further relates to diagnostic and screening methodologies
(and kits based thereon) that make use of the oncoprotein
(or antibodies specific for same) and the nucleotide
35   sequence.

SUMMARY OF THE INVENTION

It is one object of the invention to provide a

- 3 -

hepatocellular oncoprotein and a nucleotide sequence
coding for same.

It is another object of the invention to provide a
diagnostic test for the presence of hepatocellular carci-
5 nomas as well as preneoplastic or pathological conditions
of the liver.

Further objects and advantages of the present
invention will be clear to one skilled in the art from the
description that follows.

10 In one embodiment, the present invention relates
to a DNA fragment coding for the amino acid sequence set
forth in Figure 1 or an allelic variation of that se-
quence, or a unique portion thereof.

In another embodiment, the present invention
15 relates to a recombinant DNA molecule comprising:

i)   a vector, and

ii) the above-described DNA fragment.

In a further embodiment, the present invention
relates to a host cell transformed with the above-de-
20 scribed recombinant DNA molecule.

In another embodiment, the present invention
relates to a nucleotide fragment sufficiently complementa-
ry to the above-described DNA fragment to hybridize
therewith

25 In a further embodiment, the present invention
relates to a protein having the amino acid sequence set
forth in Figure 1 or an allelic variation of that se-
quence, or a unique portion thereof.

In another embodiment, the present invention
30 relates to antibodies (polyclonal and/or monoclonal)
specific for t above-described protein.

In a further embodiment, the present invention
relates to a process of producing the above-described
protein comprising culturing a host cell transformed with
35 the above-described recombinant DNA molecule under condi-
tions such that t DNA fragment is expressed and the
protein thereby produced; and isolating the protein.

In another embodiment, the present invention

- 4 -

relates to a method of detecting the presence of the above-described protein in a sample comprising:

i) contacting the sample with an antibody specific for the protein under conditions such that binding of the antibody to the protein can occur, whereby a complex is formed; and

ii) assaying for the presence of the complex.

In another embodiment, the present invention relates to a method of detecting the presence of a nucleotide sequence coding for the above-described protein in a sample comprising: i)contacting the sample with a nucleotide fragment sufficiently complementary to the nucleotide sequence to hybridize therewith under conditions such that hybridization can occur, whereby a complex is formed, and

ii) assaying for the presence of the complex.

In a further embodiment, the present invention relates to a method of diagnosing the presence of hepatocellular carcinoma in a patient comprising:

i) contacting a biological sample from the patient with the above-described antibody under conditions such that binding of the antibody to the protein present in the sample can occur, whereby a complex is formed; and

ii) assaying for the presence of the complex.

In another embodiment, the present invention relates to a method of diagnosing the presence of hepatocellular carcinoma in a patient comprising:

i) contacting nucleic acid sequences derived from a cellular sample from the patient with the above-described nucleotide fragment under conditions such that hybridization can occur, whereby a complex is formed; and

ii) assaying for the presence of the complex.

In another embodiment, the present invention relates to a diagnostic kit for detecting the presence of the above-described protein in a sample comprising a container means having disposed therewithin antibodies specific for the protein.

In a further embodiment, the present invention relates to a diagnostic kit for detecting the presence of

- 5 -

a nucleic acid sequence coding for a protein having the
amino acid sequence set forth in Figure 1 or an allelic
variation of the sequence, or a unique portion thereof,
comprising a container means having disposed therewithin
5   the above-described nucleotide fragment.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows the complete nucleotide sequence of
hhc^M, and the amino acid sequence of a 52,000 dalton pro-
tein encoded within its open reading frame.

10   Figures 2 shows the construction of hhc^M-LacZ
chimeric plasmid for the production of the hhc^M 52 kD
protein.

Figure 3 shows the Aflatoxin B$_1$ epoxide binding on
high molecular weight DNAs prepared from human hepato-
15   cellular carcinoma (Mahlavu), human normal liver and from
murine (NIH/3T3) fibroblasts.

Figure 4 shows the identification of the dG bound
by AFB$_1$ epoxide within the hhc^M (PM-1) DNA by a modified
Maxam-Gilbert sequencing method. Nucleotide sequences are
20   specified on the side. The left panel illustrates ladder
for all four 6 deoxynucleotides and AFB$_1$-dG; only native
dG and AFB$_1$- 3 were given in all other three panels on the
right. aG = AFB$_1$ bound dG at all time; °G = dG that was
not reacted with AFB$_1$; whereas °G = moderately preferred
25   dG.

Figure 5 shows the kinetic analysis of protein
production in E. coli cells harboring pJZ102. Plasmid
pJZ102 and control plasmid pJZ101 were cultured in E. coli
cells until cell density reached a Klett reading of 80, at
30   which point the inducer, IPTG (final concentration, 10^{-3}
mol), was added to activate transcription from the lac
promoter for the production of the chimeric hhc^M-lac 52-kD
protein. One ml samples of the cultures were removed at
specified times, pelleted by centrifugation and lysed, and
35   the proteins were denatured by boiling in Laemmli buffer.
Equivalent aliquots of each sample were applied and
analyzed by SDS-polyacrylamide gel electrophoresis as

- 6 -

described in (Somerville et al., Structural and Organi-
zational Aspects of Metabolic Regulation: UCLA Symposia on
Molecular and Cellular Biology, New Series, Vol. 133, p.
181-197. New York: Alan R. Liss, Inc. 1990). The lanes
5   represent: (a) pJZ102 + ITPG at time zero; (b) pJZ102 -
ITPG at time zero, and 20 hours (c); pJZ102 + ITPG at 30
minutes (d), 4 hours (e), 7 hours (f), and 20 hours (g).
Dark field microscopy of pJZ102 transformed E. coli cells
+ ITPG at 0 time (a'), 30 minutes (b'), 4 hours (e'), 7
10   hours (f'), and 20 hours (g').   Prestained molecular
weight markers (m) in kD are 130 (faint band on top), 94,
75, 50, 39, 27, 17.
        Figures 6A and 6B show purified hhc^M fusion pro-
tein p52 produced in bacteria (Figure 6A) and specificity
15   of a polyclonal anti-p52 IgG (Figure 6B). Figure 6A shows
the SDS-polyacrylamide gel electrophoresis of bacterially
expressed p52. All conditions for the bacterial expres-
sion of chimeric hhc^M-lac fusion proteins were as
described in Figure 5. Lanes d, e, and e' represent total
20   cell extracts of pJZ102-bearing E. coli cells (in varying
amounts) induced by IPTG and lane f represents the total
cell extracts f a negative control pJZ101-bearing E. coli
cells. Lane a (5µl), b (15 µl) and c (1µl) depict
different a' unts of gel purified p52 that was used to
25   immunize r/ oits. Lane m depicts pre-stained molecular
markers in kD of 75, 57, 50, 39, 27, 17.
        Figure 6B shows the reactivity of a polyclonal
anti-p52. Anti-p52 polyclonal IgG was raised by immu-
nizing rabbits. SDS polyacrylamide gel purified p52 at
30   0.8 to 1.0 mg each was used to immunize the New Zealand
White rabbit by standard techniques. Two booster injec-
tions were given. Detergent (0.2% SDS) lyzed samples
corresponding to 0.2 ml of packed human hepatoma cells
(1/3:v/v) including Mahlavu hepatocellular carcinoma,
35   Hp3p21.7 and HPG2, and pB^{r}pM-1 transfected BRL-1 tumor
cells and control BRL-1 cells and p52, at 10 µl each were
applied to sample well and allowed to diffuse and cross-
react overnight against the polyclonal anti-p52 IgG.

- 7 -

Results were recorded at 48 hours.

Figure 7 shows the DNA-DNA hybridization against ^{32}p-hhc^H DNA.

DETAILED DESCRIPTION OF THE INVENTION

5   The present invention relates to an oncoprotein coded for by a transforming nucleotide sequence of hepato-cellular carcinomas and to the transforming sequence itself. The invention further relates to unique portions (i.e., at least 5 amino acids) of the oncoprotein, and to

10   nucleotide sequences (fragments) that code for such polypeptides. The invention further relates to nucleotide segments sufficiently complementary to the above-described nucleotide sequences (fragments) to be used as probes for detecting the presence of such nucleotide sequences

15   (fragments). The invention also relates to diagnostic and screening methodologies for use in detecting the presence of hepatocel:.lar carcinomas (as well as preneoplastic or pathological conditions of the liver) in a warm blood animal.

20   The oncoprotein of the present invention is an amplified gene expression product of hepatoma cells that is specifical'v related to hepatomas. The protein can have the comj ete sequence given in Figure 1, in which case it is des gnated hhc^H. The protein can also have the

25   amino acid seq ance of a molecule having substantially the same properties (e.g., immunological) as the molecule given in Figure 1 (for example, allelic forms of the Figure 1 sequence). Alternatively, the protein (or polypeptide) of the invention can have an amino acid

30   sequence corresponding to a unique portion of the sequence given in Figure 1 (or allelic form thereof).

        The protein can be present in a substantially pure form, that is, in a form substantially free of proteins and nucleic acids with which it is normally associated in

35   the liver. The oncoprotein of the invention, including that made in cell-free extracts using corresponding mRNA, and the oncoprotein made using recombinant techniques, can be purified using protocols known in the art. The onco-

protein, or unique portion thereof, can be used as an antigen, in protocols known in the art, to produce antibodies thereto, both monoclonal and polyclonal.

5   In another embodiment, the present invention relates, as indicated above, to nucleotide sequences (fragments) (including cDNA sequences) that encode the entire amino acid sequence given in Figure 1 (the specific DNA sequence given in Figure 1 being only one example), or any unique portion thereof. Nucleotide sequences to which

10   the invention relates also include those coding for proteins (or polypeptides) having substantially the same properties (e.g., immunological) of the hhc^M polypeptide (for example, allelic forms of the amino acid sequence of Figure 1). The invention further relates to nucleotide

15   segments sufficiently complementary to the above-described nucleotide sequences (fragments) to hybridize therewith (e.g. under stringent conditions).

In another embodiment, the present invention relates to a recombinant molecule that includes a vector

20   and a nucleotide sequence (fragment) as described above (advantageously, a DNA sequence coding for the molecule shown in Figur₁ 1 or a molecule having the properties thereof). Th vector can take the form of a virus or a plasmid vecto₁ The sequence can be present in the vector

25   operably lin d to regulatory elements, including, for example, a promoter (e.g., the LacZ promoter). The recombinant molecule can be suitable for transforming procaryotic or eucaryotic cells, advantageously, protease deficient E. coli cells.

30   A specific example of a recombinant molecule of the invention is shown in Figure 2. In this example, the hcc^M nucleotide sequence is placed in a chimeric construct by replacing the codons of the original N-terminus 18 amino acids of the hhc^M p52kD with the procaryote LacZ

35   expression/translation sequence plus codons for 11 amino acids by appropriate recombinant DNA manipulations (Yang et al. Proc. of the XIV Inter. Symp. Sponsored by the International Association for Comparative Research on

Leukemia and Related Diseases Nov. 1989 (Vale, Colorado)).
Driven by the LacZ promoter, the resultant chimeric gene
is expressed at high levels in a protease deficient E.
coli mutant at 30°C.   In a further embodiment, the
5   present invention relates to a host cell transformed with
the above-described recombinant molecule. The host can be
procaryotic (for example, bacterial (advantageously E.
coli)), lower eucaryotic (i.e., fungal, including yeast)
or higher eucaryotic (i.e. mammalian, including human).
10   Transformation can be effected using methods known in the
art.   The transformed host cells can be used as a source
for the nucleotide sequence described above (which se-
quence constitutes part of the recombinant molecule).
When the recombinant molecule takes the form of an expres-
15   sion system (see specific construct described above), the
transformed cells can be used as a source for the oncopro-
tein.

The oncoprotein and nucleic acid sequence of the
present invention can be used both in a research setting
20   (for example, to facilitate an understanding of how and
why hepatocellular carcinomas develop) and in a clinical
setting to, for example, diagnosis (and/or screening) the
presence of or progress of hepatocellular carcinomas (as
well as p neoplastic or pathological condition of the
25   liver).

The diagnostic/screening methodologies referred to
above can be carried out using antisera or monoclonal
antibodies (produced using known techniques) against the
oncoprotein (or unique portions thereof) of the invention.
30   For example, the diagnostic method can take the form of an
immunoassay that can be used with urine or serum samples
of patients at high risk for hepatocellular carcinoma
(e.g. chronic hepatitis carriers) and/or of populations in
the geographically identified hot-spots of liver cancer
35   (e.g. Chitung Province of China). The screening immunoas-
say can be of the simple dip-stick type where binding of
one member of the antigen/antibody pair, attached to the
stick, with the other member of the pair, present in the

- 10 -

sample, is accompanied by a color change (such dip-stick type assays have been described for use with a variety of binding pairs). Such simple tests would be easily and widely applicable to populations in areas where analytical
5 electrophoresis equipment (required for detecting alpha-fetoprotein levels in patients' sera, which levels are currently used in screening and diagnosing the presence of hepatocellular carcinomas) may not be readily available.

The diagnostic methods of the invention can also
10 take the form of a histochemical diagnostic tests involving the use of antibodies against the protein or polypeptide of the invention. Such a test can be used on frozen or prefixed liver thin section samples to enable a more definite diagnosis of liver cancer.

15 The diagnostic methods of the invention can also involve the use of nucleic acid probes sufficiently complementary to a portion of the nucleic acid sequence of the invention to hybridize thereto. Such probes can be used to detect the presence of the endogenous sequence,
20 for example, following electrophoresis of genomic DNA digested with appropriate restriction enzymes. The probe can be labelled, for example, with $32P$, to facilitate detection.

The invention further relates to diagnos-
25 tic/screening kits for use in carrying out the above methods. The kits can comprise, for example, the above-described antibodies specific for the oncoprotein (or polypeptide) of the invention or, alternatively, the above-described nucleic acid probes, together with any
30 ancillary reagents (e.g., buffers, detectable markers, enzyme substrates, etc.) necessary to conducting the test.

The invention is described in further detail in the following non-limiting Examples.

Examples

35 The following protocols are referenced in the Examples that follow:

Molecular cloning of hhc^M

Genomic DNA purified from human normal liver and Mahlavu (African) hepatocellular carcinoma (HHC), as described below, were subjected to complete digestion by
5   HindIII restriction endonuclease.   (Other restriction endonucleases including BamHI, EcoRI and PstI, were also used for isolating genomic DNA fragments from HHC and liver DNA in an attempt to clone HHC DNA sequences; the clones isolated from these efforts were not successful
10   with respect to transfection studies.)   The DNA samples both [^{3}H]aflatoxin B$_1$ (AFB$_1$)-epoxide bound (as described below) and unbound, were separated into 180 fractions by polyacrylamide gel electrophoresis.   Specificity of [^{3}H]AFB$_1$-epoxide per µg of DNA was determined.   Fractions
15   with significant [^{3}H]AFB$_1$-epoxide specific activity were used in DNA transfection assay on NIH3T3 cells as de- scribed below. Fractions showing positive focus formation indicating positive cell transformation, were identified and the parallel unbound DNA fractions were molecularly
20   cloned by ligation onto the HindIII site of pBR322, pBR325 and/or Puc 8 plasmid DNAs for transformation of E. coli HB101 cells as described elsewhere (Yang et al., J. Gen. Virol. 19 , 63:25). Primary selection of the resultant clones was thus based on (1) the sensitivity to tetracy-
25   cline, and or color change associated with the disruption of the lacz operon containing the B-galactosidase coding sequence of the plasmid; and (2) the capability of cell- transformation in transfection assays on NIH3T3 cells with or without AFB$_1$ binding; (3) the presence of human se-
30   quence in colony-hybridization and DNA-DNA hybridization against [^{32}P]probes prepared from human Alu sequence (Lawn et al., Cell 1978, 15:1157) and also [^{32}P] labelled HindIII digested MAH HHC DNA fragments; and (4) [^{3}H] AFB$_1$-epoxide binding on the DNA fragments. After screening over 30,000
35   clones by these quadruple technical approaches including [^{3}H]AFB$_1$ binding, transfection assay on NIH3T3 cells and DNA-DNA hybridization against the [^{32}P]Alu and [^{32}P]HindIII

- 12 -

MAH HHC DNA probes, three clones were isolated.   One
particular 3.1 kb DNA restriction fragment constitutes the
hhc^M DNA.

Preparation of plasmid DNA and AFB, binding

5   The clone used in these studies has been referred
to as PM-1.   Plasmid DNA was prepared by the Holmes'
method, i.e. the rapid heating method, followed by CsCl$_2$-
ethidium bromide isopycnic centrifugation at 180,000xg for
20 hrs (Maniatis et al., Molecular Cloning, Cold Spring
10   Harbor Laboratory, Cold Spring Harbor, N.Y. 1982).   The
banded PM-1 DNA was then purified free of ethidium bromide
by isopropanol extraction and exhaustive dialysis against
TEN buffer.   A yield of 25 to 50 µg of total plasmid DNA
per 5 ml of culture was generally obtained.   The 3.1 kb
15   hhc^M DNA was then separated from PUC 8 DNA and other
contaminants by digesting the PM-1 DNA with HindIII
endonuclease and then subjecting to agarose gel electro-
phoresis and electroelution of the separated 3.1 kb band.
The resultant 3.1 kb hhc^M DNA was homogeneously purified
20   and used in AFB$_1$ activation experiments.

The hhc^M 3.1 kb DNA was also cloned into a pSVneo
vector that carried a murine retroviral (Moloney) LTR,
SV40 promoter and part of the T antigen besides the
neomycin resist nce gene.   This clone, rpMpN-1, is ex-
25   pressed at a significantly higher level when transfected
into cells and orfers special advantages for transfection
assay.

[^{3}H]AFB$_1$ at 15 Ci/mmole specificity was acquired
from Morales Laboratory, CA.   It was further purified by
30   HPLC to homogeneity and the resultant single peak of
[^{3}H]AFB$_1$ had the specific activity of 9,250 cpm/pmole.   It
was used in activation reactions with either mixed func-
tion oxidases freshly prepared from liver microsomal
preparation or by the chemical peroxidation reaction using
35   perchlorobenzoic acid and methylene chloride as described
earlier (Bennett et al., Cancer Res. 1981, 41:650; Garner
et al., Chem. Biol. Interact. 1979, 26:57).   Binding of

- 13 -

[^{3}H]AFB$_1$ epoxide with either high molecular weight HHC or plasmid DNA was monitored by kinetic analysis (Yang et al. Environmental Health Perspective 1985, 62:231 and Modali and Yang, Monitoring of Occupational Genotoxicants pp.

5   147-158 (1986)). Samples withdrawn at each time point was washed free of unbound [^{3}H]AFB$_1$ epoxide with chloroform, and ethanol precipitated prior to redissolving the [^{3}H]AFB$_1$-DNA in Tris-EDTA-NaCl (TEN) buffer for transfection assay or sequence analysis.

10   Cells, tissue culture and transfection assay

NIH/3T3 cells, passage 6 to 11, and Buffalo rat liver cells (BRL-1) for transfection assays, were maintained in Dulbecco's modified Eagle's media supplemented with 10% heat-inactivated fetal calf serum, penicillin (50

15   units ml^{-1}) and streptomycin (25 μg ml^{-1}) (DMEM) in a 5% CO$_2$ atmosphere, at 37°C.

DNA transfection was carried out as described earlier (see Yang et al. 1985 and Modali and Yang 1986, referenced above). Optimal conditions were achieved by

20   carefully titrating the pH curve for the DNA-calcium phosphate complex mixture; it was usually found that pH 6.75 ensured a fine complex precipitation.

Preparations of DNA and RNA from tissue culture cells and tumor tissues

25   Total igh molecular weight (HMW) DNA was extracted and purified from tissue culture cells and tumor tissues as described elsewhere (Yang et al., 1985 referenced above). The HMW DNA thus purified, has been subjected to proteinase K digestion, first sequential chemi-

30   cal purification with phenol-cresol, chloroform-isoamyl alcohol, ether and ethanol-NaCl precipitation, followed by RNase digestion and a second sequential chemical purification. The purified DNAs were then dialyzed against TEN buffer for use in experiments. Total RNA was extracted

35   from tissue culture cells and prepared as described previously (Maniatis et al., 1982 referenced above). Poly A rich RNA was obtained by affinity separation with oligo dT cellulose (Collaborative Research, MA.) column elution.

- 14 -

Tumorigenesis

Transformed cells, cloned out from the transfected cell culture by either cloning cylinder method or terminal dilution method, were expanded and inoculated at 10^4 to 10^6

5   cells into athymic Swiss nu/nu mice subcutaneously. Tumorigenesis in the challenged mice was monitored closely.

Nucleotide sequence analysis and site-targeted mutagenesis

Nucleotide sequencing of the hhc^M 3.1 kb and

10   variants produced by site-targeted mutagenesis were carried out by the standard Maxam-Gilbert Methods in Enzymology 1980, 65:499 and the Sanger (M13) dideoxy sequencing methods (Maniatis et al., 1982 referenced above).

15   Specified oligonucleotide sequence of 20 mers carrying the targeted dG--->T mutation were synthesized by the Applied Biosystem oligonucleotide synthesizers. They were used as templates in generating the mutated clones. Mutant DNA clones were produced in accordance with the

20   protocol provided by and using the oligonucleotide-directed in vitro system of Amersham (Arlington Hts., IL). DNAs of the mutat i clones were verified by nucleotide sequencing. Effect of these site-targeted mutagenized DNA were analyzed by potentiation of cell-transformation in trans-

25   fection assa / on NIH/3T3 cells and RNA expressions in transfected cells using the BRL dot-blot technique (Bethesda Research Laboratory, Rockville, MD).

Example I

Dosimetry of AFB₁ binding and potentiation of hhc^M
cell-transformation capability on NIH/3T3 cells

30   AFB₁ epoxide binds high molecular weight DNAs prepared from human hepatoma, human liver and mouse NIH/3T3 cells efficiently (Fig. 3). The initial rates in each binding kinetic were extremely rapid. The rates of

35   AFB₁-epoxide binding to human normal liver or hepatoma DNA and to murine NIH/3T3 cell DNA became significantly different after one minute of binding reaction. The MAH HHC DNA showed a greater rate of binding than normal liver

- 15 -

DNA and all the dG targets became saturated earlier, whereas AFB_1 epoxide bound the normal liver DNA at a slower rate but eventually saturated all the dG targets at a slightly lower level. The human DNAs showed a higher

5 level of AFB_1 binding than the murine NIH/3T3 cell DNA. The overall AFB_1 specific activity, i.e. AFB_1-dG adduct, was found to be about one dG bound per 10 nucleotides among these high molecular weight double stranded DNAs. This overall specificity also took into consideration the

10 existence of secondary or tertiary structure of the high molecular weight DNAs. AFB_1 epoxide binding on linearized 3.1 kb double stranded hhc^M DNA was consistently found to be 4 to 8 dG bound per 10^4 nucleotides. This higher binding capability reflects the relatively easy accessi-

15 bility of dG within the linearized double stranded PM-1 DNA by AFB_1 epoxide and should not be compared with the efficiency of AFB_1-dG adduct formation with high molecular weight native double-stranded DNA.

Within a finite dosimetry the binding of AFB_1

20 epoxide with dG potentiates the cell-transformation capability of hhc^M by 10 to 20 fold as seen in the experiment illustrated in Table 1.

- 16 -

Table 1.   AFB₁ Dose-dependent Activation of PM-1 DNA in Transformation of NUB/3T3 Cells

DNA Source	AFTB₁ femtomole per 100 ng DNA	Number of Foci per 100 ng DNA
hhc^M (PM-1)	0	15×10^{-1}
C-Ha-ras-1	0	465
c-K-ras-1	0	0
c-hhc (human liver homolog)	0	0
E. coli	0	0
hhc^M (PM-1)	0	15×10^{-1}
hhc^M (PM-1)	5	18
	14	26
	24	66
	35	3
c-hhc	0	0
	8	0
	15	0
	30	0
	40	0

AFB₁ binding and transfection assay were as described in Methods. Data were calculated on the basis of per 100 ng. In the assay with unbound hhc^M DNA the transfection assays were carried out with 500 ng to 1.5 ug of DNA in order to obtain reasonable foci formation on NIH/3T3 cells. Transfection with AFB₁-epoxide bound DNA was carried out at a range of 50 to 500 ng DNA. Data were normalized to show potentiation of hhc^M cell-transformation capability by AFB₁-epoxide activation.

Whereas the efficiency of unbound PM-1 DNA in transforming NIH/3T3 cells was usually observed at about 15 FFU/µg DNA the efficiency of AFB₁ epoxide activated PM-1 DNA was optimized at 66 FFU/100 ng DNA, an increase of 20 fold. The possibility of non-specific mutagenization accounting for this potentiation were considered. That this potentiation effect was due to free AFB₁ that diffused into the cell or recycling of AFB₁ adducts has been ruled out earlier with the appropriate control experiments which showed that activation of normal liver or E. coli DNA at the same dosimetry failed to activate any cell-transforming capability (Yang et al., 1985 referenced above).

Moreover in this experiment with AFB_1 activated DNA from c-ras^k-1 or c-hhc, a normal human liver homolog to hhc^M as the appropriate controls, no cell-transformation of NIH/3T3 cells was obtained suggesting that AFB_1 epoxide

5   activated PM-1 DNA was not a random phenomenon. Moreover the AFB_1 dose dependency of PM-1 DNA in cell-transformation efficiency (Table 1) further substantiated the specificity of AFB_1 epoxide binding in conferring the potentiation of cell-transformation. Whereas optimal

10   dosimetry was seen at 24 femtomole AFB_1/100 ng of PM-1 DNA, at dosimetry beyond 45 femtomole per 100 ng of PM-1 DNA, an overkill effect was observed. No transformed foci were obtained in NIH/3T3 cells transfected with AFB_1 epoxide bound PM-1 DNA although human DNA was incorporated

15   into the NIH/3T3 cells in a degraded form (Yang et al., 1985 and Modali and Yang, 1986 referenced above). This observation suggested that over activation of PM-1 DNA not only generated scissions in the molecule but possibly degradation leading to a loss of biological activity. It

20   was also evident from these results that no more than one or at most few AFB_1-dG adducts per PM-1 DNA molecule uld be tolerated by the hhc^M DNA before the biological activity of the hhc^M DNA became compromised and at the risk of survival. Moreover the potentiation of hhc^M DNA in

25   cell-transformation probably necessitates no more than one or at most a finite number of AFB_1 bindings.

EXAMPLE II

Specificity of the AFB_1-epoxide binding on dG's of PM-1 DNA

30   Deoxyguanine nucleotide of native DNA, when bound by AFB_1 epoxide, became alkali and therefore could be identified by piperidine cleavage; whereas unbound deoxyguanine nucleotide within the same native DNA would not cleave without dimethyl sulfide (alkali) treatment.

35   Figure 4 shows the dG targets within the PM-1 DNA when bound at a saturation conditions. When the targeted sequences are evaluated in sets of tetranucleotides, an empirical formulation can be derived on the basis of the

- 18 -

binding pattern of AFB$_1$ epoxide with the dG's in PM-1 DNA.
Table 2 summarizes the nucleotide sequence in a set of
tetranucleotides that were seen and targeted by AFB$_1$
epoxide. As shown in Figure 4, the dG within a sequence
5   of any one of the following tetranucleotides of AGAG,
AGTT, TGTT, TGAT, or AGAA, escaped attack by AFB$_1$ epoxide
and hence showed no cleavage in the sequence without prior
DMS treatment. This is confirmed by the distinct cleavage
of dG as a result of AFB$_1$ epoxide attack on dG in a se-
10   quences of GGGC, CGGC, AGGC, TGGC or CGCG. Upon evaluat-
ing the various sequences in which a dG target could be
accessed by AFB$_1$ epoxide, it can be concluded that within
a double stranded DNA, the least likely dG target would be
that flanked by dA and dT, i.e. category III. The most
15   likely dG target would be that flanked by dG and/or dC,
i.e. category I, and that tetranucleotide sequences in
which dG is either preceded by dA or T and followed by dG
and dC would be the moderately preferred targets of AFB$_1$
epoxide, i.e. category II. This, of course, does not take
20   into consideration the secondary or the tertiary structure
of the DNA in its natural state since these analyses were
done on linearized double-stranded DNA. It should also be
mentioned that whereas the dG binding affinity of
AFB$_1$-epoxide was greatly affected by the vicinal nucleo-
25   tides in the double-stranded PM-1 DNA, no specificity was
observed with respect to AFB$_1$-epoxide binding to dG in
single stranded DNA. The observations of Modali and Yang
(1986 referenced above) were basically in agreement with
others working on AFB$_1$ binding on OX174 and pBR 322 DNAs
30   (Misra et al., Biochemistry, 1983, 22:3351).

Within the past two years, the nucleotide sequence
of hhc^M has been resolved by a combination of
Maxam-Gilbert nucleotide sequencing technique and the M13
dideoxy method using the BRL kilobase sequencing system.
35   Applying these empirical rules in computer analysis of the
hhc^M 3.1 kb nucleotide sequence, the most and moderately
preferred dG targets within the various loci of hhc^M have

been predicted (Table 3). Although a maximum number of 60
dG targets was predicted on the basis of AFB_1-epoxide
binding studies with linearized 3.1 kb hhc^M DNA, it was
evident upon examining the possible secondary and tertiary
5   structure of hhc^M sequence, that a much lower number of dG
targets would be accessible by AFB_1-epoxide.   Moreover,
only a few such induced mutations would produce any effect
of survival value.

- 20 -

Table 2. Vicinal Nucleotide Sequence Dictates the dG
          Targets of AFB₁-Epoxide Binding*

Preferred targets Category I	Least Favored Targets Category III
*	*
GGGG	AGAG
GGGC	AGTG
GGGA	AGAA
GGGT	AGAC
	AGAT
CGGG	TGAG
AGGG	TGAC
TGGG	TGAA
	TGAC
CGGC	TGTG
AGGC	TGTA
TGGC	TGTC
	TGTT
CGGA	
AGGA	
TGGA	
CGGT	
AGGT	
TGGT	

*This table represents the dG targets of AFB₁-epoxide
birding observed in studies with linearized double strand-
e⁻ PM-1 DNA. Moderately preferred dG targets, i.e.
tegory II, are omitted here but are described elsewhere
Modali and Yang, 1986).

Table 3. <u>Predicted dG Targets within the Nucleotide Sequence of hhc^H Preferentially Attacked by AFB$_1$-Epoxide</u>

CGCC	CGGC	GGCC	GGGC	GGGA	AGGA	TGCC	TGCG	TGGA	TGGG	GGAC
	CGGC			GGGA	AGGA			TGGA	TGGG	
					73					
					74					
						84				
								97		
								98		
				125						
				126						
									140	
								221		
					223					
					224					
					307					
					308					
								371		
								391		
			472							
									481	
									492	
				494						
				495						
					539					
										550
				560						
				561						
			577							
				692						
									860	
					901					
					1125					
				1320						
				1321						
								1330		
									1354	
									1404	
		1405								
				1431						
		1543								
		1588								
									1637	
						1652				
				1765						
					1815					
		1853								
					1862					
										1868
							1878			
					1986					

- 22 -

Table 3. (con't)

	CGCC	CGGC	GGCC	GGGC	GGGA	AGGA	TGCC	TGCG	TGGA	TGGG	GGAC
		*			*	*			*	*	
		CGGC			GGGA	AGGA			TGGA	TGGG	
					2064						
						2094					
			2205								
									2315		
									2331		
				2352							
				2352							
									2460		
			2482								
									2718		
								2797			
									2884		
								2926			

In order to analyze the possible effect of any such AFB_1 induced dG-->T mutation, site-targeted mutagenesis study of the hhc^M DNA was initiated using polynucleo-

5   tides of 20 mers that carried a predicted dG-->dT point-mutation, presumably the result of an AFB_1-epoxide mutagenesis. Thus far, only a few of the predicted dG--->dT mutagenesis sites have been analyzed and these are summarized in Table 4. The recombinant construct carrying

10   the hhc^M sequence in the SV40 T antigen vector plus a neomycin resistance marker, rpN^{r}pM-1 was used in this study since it offered the advantage of selecting the transfected cells by its resistance to Gentamicin sulfate (G418), an analog of neomycin. Using expression of hhc^M

15   specific mRNA as a criterion, we analyzed by Northern dot-blot in a semi-quantitative assay of the mRNA, i.e. poly A enriched RNA, expressed in the G418 resistant NIH/3T3 cells after transfection with the mutagenized hhc^M sequence. Focal transformation in these cells was moni-

20   tored for 4 to 6 weeks.

Results from seven mutagenized clones, for which nucleotide sequence confirmation was available, suggested that, thus far, mutation leading to a structural protein alteration did not seem to potentiate the

25   cell-transformation of hhc^M (Table 4). Alternatively the introduced dG-->T mutations which led to amino acid substitution, thus far, have not altered cell-transformation or expression of mRNA levels. These included mutation at 577 which caused an amino acid substitution of

30   Gly--->Val, and mutation at 1005 which resulted in no amino acid substitution because of the wobbling code.

Within the hhc^M nucleotide sequence, there exists an apparent open reading frame, ORF, coding for a polypeptide of about 467 amino acids. This was in good agreement

35   with a 55-57 kD protein and some smaller polypeptide including one 53 kD protein observed in cell-free protein synthesis using hhc^M-specific mRNA in a rabbit reticulo-

- 24 -

cyte lysate system. dG--->T mutations at nucleotide 73
and 74 in the 5' terminus, which bears the consensus
sequence for ribosomal RNA binding site just 5' ahead of
the first methionine codon, blocked cell transformation
5   although hhc^M specific mRNA level showed no difference.
This could be the result of blocking protein synthesis.
Likewise, interpreted as a mutations at 492 and 550 also
blocked cell-transformation since a stop codon (UGA) was
introduced in each case to stop protein synthesis
10   prematurely.

     It was of interest to note that dG-->T mutation at
626 generated a sequence resembling the enhancer sequence
for RNA polymerase II, which was reported to function even
within the coding sequence (footnote of Table 4). The
15   level of mRNA level was increased by 1.5 fold and cell
transformation seemed to be enhanced by a slight increase
in the number of foci per μg of DNA. This observation
suggested that one possible action by which AFB$_1$ induced
mutation in hhc^M, which itself is a moderately transforming
20   DNA sequence, led to increase in its transformation
potential is through augmentation of hhc^M expression.
This is analogous to other observations which also indi-
cated that an elevated expression of the cellular ras
proto-onocgene driven by a murine LTR sequence, containing
25   both promoter and enhancer sequence, also led to cell
transformation in tissue culture cells predisposed to
immortality.

Table 4.   The Effect of dG --> dT Mutation Induced by
           Site-Targetted Mutagenesis Within The hhc^M DNA
           Sequence

# on hhc^M	Sequence	mRNA Synthesis[1]	Cell Transformation[1]
73	AGGA --> ATGA	+	_@1
74	AGGA --> AGTG	+	_@1
492	TGGG --> TGTG	+	-@2
550	GGAG --> GTAG	+	-@2
577	GGGC --> GTGC	+	+
526	GGGG --> GTGG	++	..@3
1005	TGCA --> TTCA	+	+

@1   Disruption of ribosomal RNA (16S) binding site: AGGA.

@2   Creation of stop codon: UGA.

@3   Creation of an enhancer sequence: GGTGTGGTAAAG
     (Watson et al., 1987; Dynan and Tjian, 1985; Schaffner
     et al. 1985) and hence increases expression.

     Cell transformation was determined by transfection
     analysis as described in Methods and mRNA synthesis in
     transfected cells was determined by Northern dot-blot
     analysis with [^{32}p}3.1 kb hhc^M DNA.

- 26 -.

Example III

Hhc^M-p52 and anti-p52 and their use as screening and diagnostic reagents for human hepatocellular carcinoma and related liver preneoplastic pathological conditions

5   Hhc^M-p52 as a fusion protein was produced by a bacterial system described above at high levels (Figure 5). This protein was used to generate a panel of both monoclonal and polyclonal antibodies against related human hepatoma proteins (see Figures 6A and 6B). Anti-p52, a

10   polyvalent antibody against hhc^M-p52 was produced and shown to be highly specific against an African (Mahlavu) hepatoma and a Philadelphia hepatoma (Figures 6A and 6B).

       Assays for the presence of hepatoma specific

15   protein p52 in tumor samples entail diffusion and immuno-precipitation using the tumor sample extracts reacted with anti-p52, with or without radioactive or immunofluores-cence labels. Further, anti-p52, labelled with either a radioactive compound or with a chromophore, is useful in

20   RIPA or colorchange assays, respectively, for testing for the presence of hepatoma related proteins shed by the patient in sera and urine samples. Fluorescence imagery analysis using anti-p52 conjugated to a fluorescence compound or another suitable compound for systemic perfu-

25   sion, provide the ability to localize in situ preneo-plastic or neoplastic lesions by scanning. Localization of lesions permits laser removal with surgical precision, and/or other treatment.

       Hhc^M-p52 nucleotide sequence, labelled appropri-

30   ately, can be applied to diagnose hepatomas in biopsy samples. Hhc^M-related nucleic acid sequences can be detected in needle biopsy samples of patients suspected of carrying preneoplastic nodules or liver cancer. This is accomplished by the using the polymerase chain reaction to

35   amplify "hhc^M-like" sequences using fragments of the hhc^M-p52 sequence as primers, and then detecting the presence of such hhc^M-like sequences in the biopsy sample with labelled hhc^M-p52 as a probe in a DNA-DNA hybridization

- 27 -

reaction.   Such an example is shown in Figure 7.

*   *   *   *

The entire contents of all references cited herein are hereby incorporated by reference.

5   The present invention has been described in some detail for purposes of clarity and understanding.   One skilled in the art will appreciate from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the inven-

10   tion.

- 28 -

WHAT IS CLAIMED IS:

1.   A DNA fragment coding for the amino acid sequence set forth in Figure 1 or an allelic variation of said sequence, or a unique portion thereof.

5   2.   The DNA fragment according to claim 1 wherein said fragment codes for the amino acid sequence set forth in Figure 1, or a unique portion thereof.

3.   A recombinant DNA molecule comprising:

   i)   a vector, and

10   ii)   said DNA fragment according to claim 1.

4.   The recombinant molecule according to claim 3 wherein said DNA fragment codes for the amino acid sequence set forth in Figure 1, or a unique portion thereof.

5.   The recombinant DNA molecule according to
15   claim 3 further comprising a promoter sequence operably linked to said DNA fragment.

6.   A host cell transformed with the recombinant DNA molecule according to claim 5.

7.   The host cell according to claim 6, wherein
20   said cell is a procaryotic cell.

8.   The host cell according to claim 7, wherein said cell is an E. coli cell.

9.   A nucleotide fragment sufficiently complementary to said DNA fragment according to claim 1 to hybrid-
25   ize therewith.

10. A protein having the amino acid sequence set forth in Figure 1 or an allelic variation of said sequence, or a unique portion thereof.

11. The protein according to claim 10 wherein
30   said protein has the amino acid sequence set forth in Figure 1, or a unique portion thereof.

12. Antibodies specific for said protein according to claim 10.

13. The antibodies according to claim 12, wherein
35   said antibodies are polyclonal.

14. A process of producing the protein according to claim 10 comprising

           culturing a host cell transformed with a

recombinant DNA molecule comprising:

       i)   a vector, and

       ii)   a DNA fragment coding for said protein under conditions such that said DNA fragment is expressed and said protein thereby produced; and

       isolating said protein.

16.   15. A method of detecting the presence of the protein according to claim 10 in a sample comprising:

       i)   contacting the sample with an antibody specific for said protein under conditions such that binding of said antibody to said protein can occur, whereby a complex is formed; and

       ii) assaying for the presence of said complex.

16. The method according the claim 15 wherein said antibody is linked to a detectable label.

17. A method of detecting the presence of a nucleotide sequence coding for said protein according to claim 10 in a sample comprising:

       i)   contacting the sample with a nucleotide fragment sufficiently complementary to said nucleotide sequence to hybridize therewith under conditions such that hybridization can occur, whereby a complex is formed, and

       ii) assaying for the presence of said complex.

18. The method according to claim 17 wherein said nucleotide fragment is linked to a detectable label.

19. A method of diagnosing the presence of hepatocellular carcinoma in a patient comprising:

       i)   contacting a biological sample from said patient with an antibody according to claim 12 under conditions such that binding of said antibody to said protein present in said sample can occur, whereby a complex is formed; and

       ii) assaying for the presence of said complex.

20. The method according to claim 19 wherein said sample is a tissue sample.

- 30 -

21. A method of diagnosing the presence of
hepatocellular carcinoma in a patient comprising:

i)   contacting   nucleic   acid   sequences
derived from a cellular sample from said patient with said
nucleotide fragment according to claim 9 under conditions
such that hybridization can occur, whereby a complex is
formed; and

ii)   assaying   for   the   presence   of   said
complex.

22. A diagnostic kit for detecting the presence
of the protein according to claim 10 in a sample compris-
ing a container means having disposed therewithin antibod-
ies specific for said protein.

23. A diagnostic kit for detecting the presence
of a nucleic acid sequence coding for a protein having the
amino acid sequence set forth in Figure 1 or an allelic
variation of said sequence, or a unique portion thereof,
comprising a container means having disposed therewithin
said nucleotide fragment according to claim 9.

Q12559

```
                                                   54                                                    81
AAG CTT AAT AGA AAA TAT GAG CAA CAT ACA CAA ACA TTA GCA ACA ATG [ATA TAA] AAT ACC ACT TAA ACA TAA GGA AAA [ATG]
                                                                                                          [MET]

                          108                                 135                                      162
TTG CCC TTC ACT TGT GGA AGA AAT GCA AAT GAA AAC AGC CCT AGG GAT GTT GGG GTG GCA CCT GCT GCA GAG GGT
Leu Pro Phe Thr Cys Gly Arg Asn Ala Asn Glu Asn Ser Pro Arg Asp Val Gly Val Ala Pro Ala Ala Glu Gly

                          189                                 216                                      243
AAC GTG CAG CAT GTC GAG AGC GGC ACT GCC AAG GAG ACT TTG AGC TCA AGG TCA GGA GGA GGT AGT CTC TCC CAT CTC
Asn Val Gln His Val Glu Ser Gly Thr Ala Lys Glu Thr Leu Ser Ser Arg Ser Gly Gly Gly Ser Leu Ser His Leu

                          270                                 297                                      324
TTC GAG TGC AGC TCT AAA CCC TGC AAA CAC CTG GAG AAG CTA TCT GAG CTG CCT CCA GGA LA` ATG CAA ATG GAC
Phe Glu Cys Ser Ser Lys Pro Cys Lys His Val Leu Glu Lys Glu Leu Pro Pro Gly His MET Gln MET Asp

                          351                                 378                                      405
ACT CTG ATC ATA AAA TTA TCA AGA TTG AGA AAT AAG ACA AAA ATG GAG GTG CCA CCA AAC CAG TGG AAA TTT TTC CCC
Thr Leu Ile Ile Lys Leu Ser Leu Arg Leu Arg Asn Lys Thr Lys MET Glu Val Pro Pro Asn Gln Trp Lys Phe Phe Pro

                          432                                 459                                      486
TTT TCA TTC CTC TGG CAT TCC CTG GCC TTG ACT CAA GGC AGC CCA CAC TCT AGG AGA CAC CAG GGC ACA GGT GGG GAG
Phe Ser Phe Leu Trp His Ser Leu Ala Leu Thr Gln Gly Ser Pro His Ser Arg Ser Arg His Gln Gly Thr Gly Gly Glu

                          513                                 540                                      567
CTC TGG GGG ACC CTC CAG GCT TAC TCA GTG AAT GGG TTA GCA GCA GCC ACA GGA GCC ATG GAG CCT GCA GGG ACC CAC
Leu Trp Gly Thr Leu Gln Ala Tyr Ser Val Asn Gly Leu Ala Ala Ala Thr Gly Ala MET Glu Pro Ala Gly Thr His
```

FIG. 1

```
594                                                                                          648
AAC ACT GAG GGC AGG GAT CTT GCC TCT AAT CAG ATA AGC TGT GAT TCC CGA GAG GTA AAG GCC ACG GGT CTT TTT
Asn Thr Glu Gly Arg Asp Leu Ala Ser Asn Gln Ile Ser Cys Asp Ser Arg Glu Gly Val Lys Ala Thr Gly Leu Phe

675                                                                                          729
CTC TCC ACA TCT TCC CAC GTC ATG ACC CCA GAG AGA GGG AGA AAG TGT GAG CAC CGT GAC ATA ATG AGC CGC AGC
Leu Ser Thr Ser Ser His Val MET Thr Pro Glu Arg Gly Arg Lys Cys Glu His Arg Asp Ile MET Ser Arg Ser

756                                                                                              810
CTT CTG ACT AGA TGC CCC AAA GAA TCC CAG GTG ACC ACA CAG CAT CAG AGA AAC TGC AGG GTA ATG AGG AAC TTT GGA
Leu Leu Thr Arg Cys Pro Lys Glu Ser Gln Val Thr Thr Gln His Gln Arg Asn Cys Arg Val MET Arg Asn Phe Gly

837                                                                                                  891
AAG CAA TCC ATC GTG TTG TCA GTA AAA CCT CTG GCT CAC TCC CGA GCT GGG CAT GCA TGG ATG GTG ACC CTC GAT GGA ATA
Lys Gln Ser Ile Val Leu Ser Val Lys Pro Leu Ala His Ser Arg Ala Gly His Ala Trp MET Val Thr Leu Asp Gly Ile

918                                                                                              974
GAC TAT GAG GAA CCA GGT GAG GGG ATC TAC CTC CAC CGA GAC GTG AGA GTG ACC TGC ATA CCC AAA CAC CAT GAG GCT TTA
Asp Tyr Glu Glu Pro Gly Glu Gly Ile Tyr Leu His Arg Asp Val Arg Val Thr Cys Ile Pro Lys His His Glu Ala Leu

999                                                                                              1053
AAG ACT GAG CTG ATG TGG AAG CCA CAG CCT CTG CAG GTT GCT CTG CAC TTG CAA CAT AAG CCC AAC CAC ATC AAT TGC TGC
Lys Thr Glu Leu MET Trp Lys Pro Gln Pro Leu Gln Val Ala Leu His Leu Gln His Lys Pro Asn His Ile Asn Cys Cys

1080                                                                                                 1134
AAA ACA AAA CTA CAG CAT TCT CCA TAC CAC TTA AAT AAG ACA CAG AGT CTC ACA ACA TTC AAA ACG CCC AGG ACA CAA TCC
Lys Thr Lys Leu Gln His Ser Pro Tyr His Leu Asn Lys Thr Thr Gln Ser Leu Thr Thr Phe Lys Thr Pro Arg Thr Gln Ser
```

FIG. 1(cont.)

```
                                          1188                              1215
     1161
AAA ATT ACT TCT ACA AAA AAT CAG GAA AAT CTC AAT GAG CAA GGA AAA TGG CAA TCA GTA GCT GCC AGT GCT GAG ATG ACA
Lys Ile Thr Ser Thr Lys Asn Gln Glu Asn Leu Asn Glu Gln Gly Lys Trp Gln Ser Val Ala Ala Ser Ala Glu MET Thr

                                          1270                              1296
     1248
ATG AGG GTT GGA ATC ATC AAC ATC TTT AAA GTA ATT ATC ATA AGC ATT CTC CAG CAA GTA ATG GCA AAC ACT CTT GAG ATA
MET Arg Val Gly Ile Ile Asn Ile Phe Lys Val Ile Ile Ile Ser Ile Leu Gln Gln Val MET Ala Asn Thr Leu Glu Ile

                                          1350                              1377
     1323
AAT GGA AAG ATA AGA AGG CTC AGG GAG AAA GTG GAA TGT ACA AAG AAT GAC CAA CAA GTT GCA CCA CTG GAA ACA AAT
Asn Gly Lys Ile Arg Arg Leu Arg Glu Lys Val Glu Cys Thr Lys Asn Asp Gln Gln Val Ala Pro Leu Glu Thr Asn

                                          1431                              1458
     1404
CAC CAG GAT AAA GCA GTC TCT GGC TGG GCC AAC AGG AGA ATG GAA GTT GTT ATG GCA GTT GTC CAA
His Gln Asp Lys Ala Val Ser Gly Trp Ala Asn Arg Arg MET Glu Val Val MET Ala Val Val Gln

                                          1512                              1539
     1485
TTT GAA CAA CAC AAA AGA CAC [TGA] GGC AGG GCT CAG TGG CTC ACA CC[T ATA A]TC CCA ATA CCT TGG
Phe Glu Gln His Lys Arg His  ***  Gly Arg Ala Gln Trp Leu Thr Pro Ile Ile Pro Ile Pro Trp

                                          1593                              1620
     1566
GAG GCC GAG GCA [ATG] TAT CAC CTG AGG TCA GGA GTT CAA GAC TAC CCT GGC CAA CAT GGC AAA ATC CCA TCT CTA CTG AAA
Glu Ala Glu Ala [MET] Tyr His Leu Arg Ser Gly Val Gln Asp Tyr Pro Gly Gln His Gly Lys Ile Pro Ser Leu Leu Lys

                                          1674                              1701
     1647
ATA CAA GAA TTA GCT GGG CAT GGT GGC AGG TGC CTG CAA TCC CAG CTA CTC AGG AGG CAG GAG AAT CAC TTG AAC
Ile Gln Glu Leu Ala Gly His Gly Gly Arg Cys Leu Gln Ser Gln Leu Leu Arg Arg Gln Glu Asn His Leu Asn
```

FIG. 1(cont.)

1728
TCG GGA GGT AGA GGG TGC AGT GAG CCA AAA TCG CAC CTC TGC ATT CCA GCC TGG GTG ACA GAG GGA GAC TCT GTC TCA AAA 1782
Ser Gly Gly Arg Gly Cys Ser Glu Pro Lys Ser His Leu Cys Ile Pro Ala Trp Val Thr Glu Gly Asp Ser Val Ser Lys

1809
CAA AAC ACA AAA AAT GAA CAG CAC CTC AGG AAC AAT ACC AAA AAG TCC AAC AGC TGT ATA ATT GGT GGC CCA GAA GGA 1863
Gln Asn Thr Lys Asn Glu Gln His Leu Arg Asn Asn Thr Lys Lys Ser Asn Ser Cys Ile Ile Gly Gly Pro Glu Gly

1890
GAG GAG AAA GAG TGG AGT ACA GAA ATG AGA TCT GAA GAA CTA ATG ACT GTT AAT AAG GAC ATA AAC 1944
Glu Glu Lys Glu Trp Ser Thr Glu MET Arg Ser Glu Glu Leu MET Thr Val Asn Lys Asp Ile Asn

1971
CTA AAG ATA GAT TCA AAA GCC CAG CTG AAT TCA AAT AGG ATA AAT ACA GAT GCA GAT ATA TTA TCA TTA AAC TGT GAA 2024
Leu Lys Ile Asp Ser Lys Ala Gln Leu Asn Ser Asn Arg Ile Asn Thr Asp Ala Asp Ile Leu Ser Leu Asn Cys Glu

2052
ATA AAT TGG TTT TGT CAC AAG CCA TTG TCA CTG TGG GAG AAA AGA TAC CAA AAG GAA GGA AAT ACA 2106
Ile Asn Trp Phe Cys His Lys Pro Leu Ser Leu Trp Glu Lys Arg Tyr Gln Lys Glu Gly Asn Thr

2133
GAA TAT TAT GGC CAT GGG AAA GAG GTG TCA GTG TGA ATA CAT AGA ACA GCA CTT AGC AAC AAC CCC AAA TGA TGG GGC 2187
Glu Tyr Tyr Gly His Gly Lys Glu Val Ser Val • Ile His Arg Thr Ala His Leu Ser Asn Asn Pro Lys • Trp Gly

2214
TTC CTA CAA AAC AGT TGG CCT TTA CTC TTC AAA AGT GTC AGG TCA CGA AAT AAA TCC ATG CTG AGG ACC CGT TCC AGG TTA 2268
Phe Leu Gln Asn Ser Trp Pro Leu Leu Phe Lys Ser Val Arg Ser Arg Asn Lys Ser MET Leu Arg Thr Arg Ser Arg Leu

FIG. I (cont.)

```
2295                                                    2322                                              2349
AAG CAG ACT AAA GGG GCT GGA CAA CAC AGT GAA CAC TGT GAG CTT GGA TTA GAT ATA TGC TGG ACT AGA GAA GGC TGT GAG
Lys Gln Thr Lys Gly Ala Gly Gln Gln Pro Ser Glu Thr Cys Glu Leu Gly Leu Asp

2376                                                    2457                                  2484
GGG GAC AAT GGC TGA AAT GTG ACT ATA GTT ACT ATG AAA AAT GTT AGG ACT TGG AAA ATC TAT ATA AAG CAG ACG

2511                                                    2538                                              2565
GCA TAA TTC TTG TAC TTT TTT TGC AAC TTT TTA ATA AAC CTG AAA CTA TTT CAA AAT GAA AAG TTA ATC CAA GCT GTC TTG

2592                                                    2619                                          2646
AGT AGA AGT TAA AAC AAC AAC AAA AGA AAA TTG AAA AGT TAA AAA TGA ACC CCC AAC AGA ATG TTC CCC TTT ATT TTT

2673                                                    2700                                              2727
CTT TCA TGT AAG GAC GCA GGA TAT GCA TTT TGC TCA GCT ACC ACC CTT CAC TGC ATC CCA TTT TGA GAA GTG GTA TTT TCT

2754                                                    2781                                          2808
TCA TTC ATC TGT TCT AGG TTT TTA AAA TAT TTA AGA TCT TCT CTT TTT AAA GAA TCT GTT CAT TTG GAA TGT ACT TTT

2835                                                    2862                                              2889
TGC ATT TTT ACT TGT GAA AAT ATG TAT TTA TCC TTT TTG TTA TGA ATG TAT GAC TTC ACT GTG TCA GAG AAT ATG GTC TTA

2916                                                    2943
AGA GAT ACA GAA AAC TTT TGA GAA TGA TAA GAT CTG GAC ATG CTA GAT GAA ATC AAA GCC CTG GAT GTC CTT GTT CAA GCT T
```
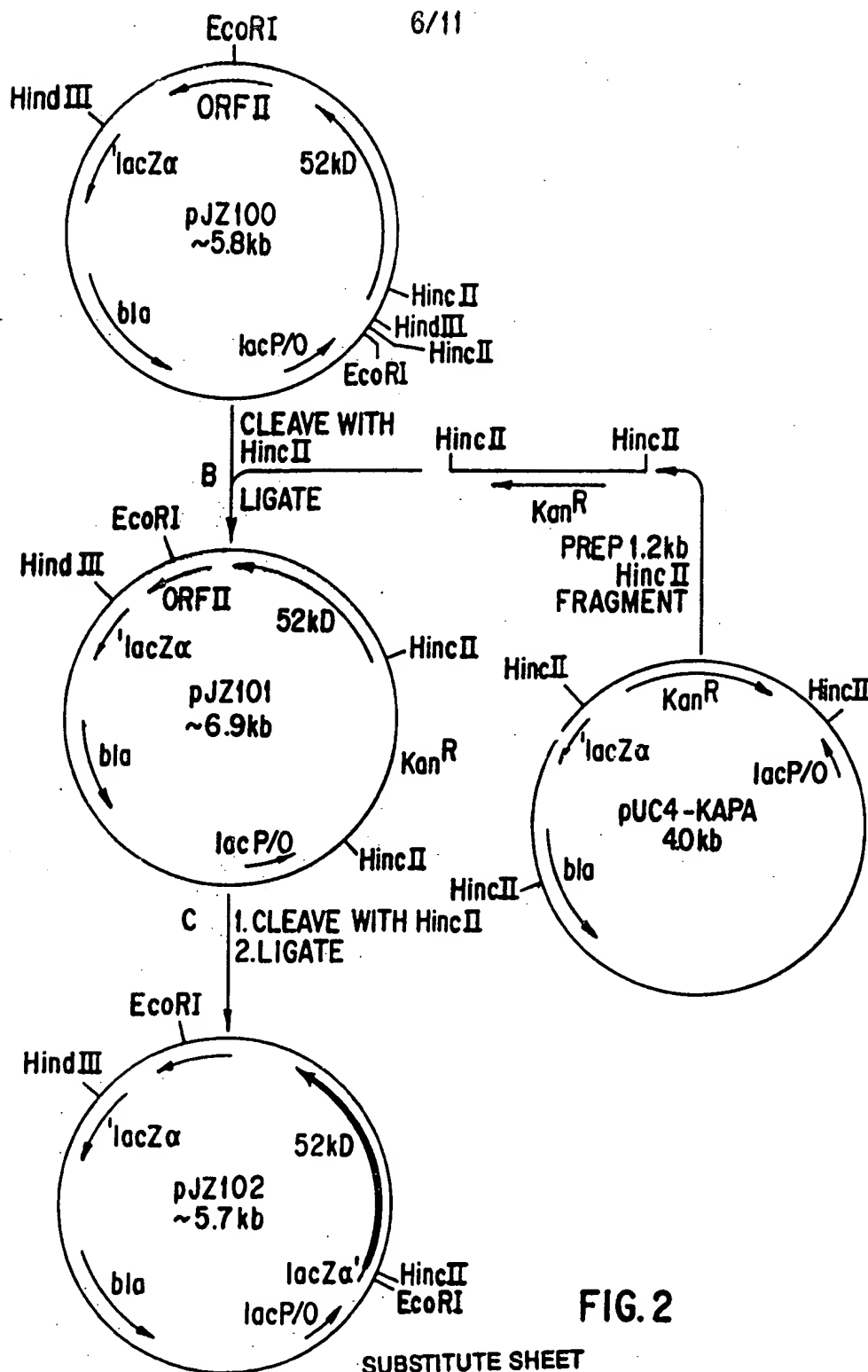
FIG.1 (cont.)

FIG. 2

SUBSTITUTE SHEET

7/11

FIG. 3

8/11

FIG. 4

9/11

FIG. 5

10/11

FIG. 6B

FIG. 6A

11/11

FIG. 7

CONTROL

PCR AMPLIFIED
HEPATOMA
DNA SAMPLE
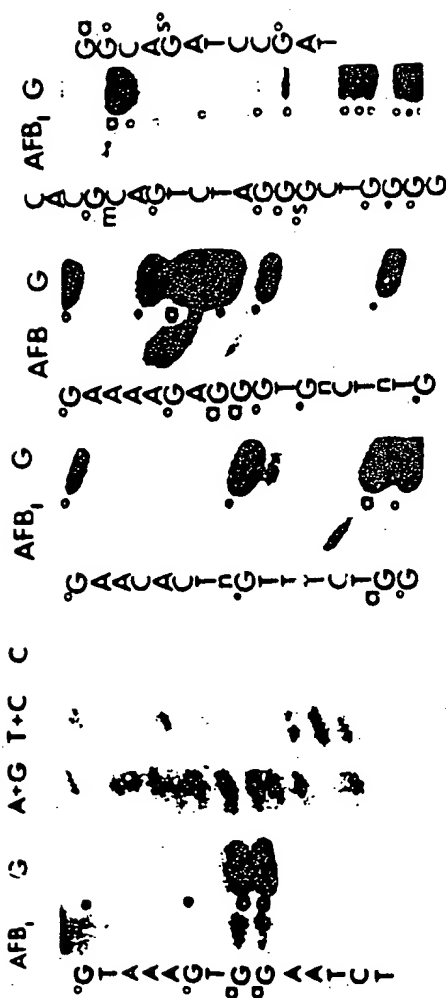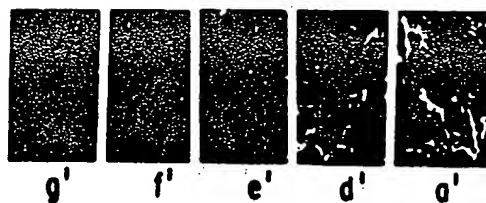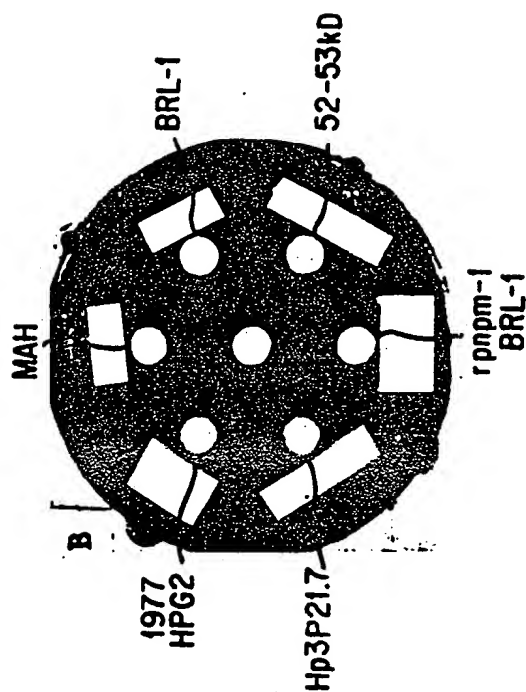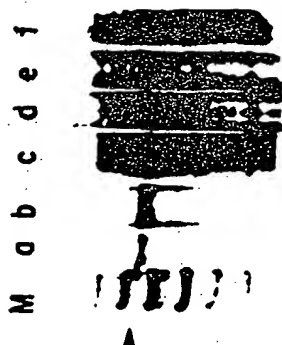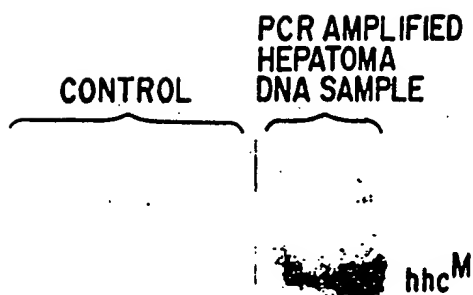
hhc^M

DNA-DNA HYBRIDIZATION
AGAINST (^{32}P)-hhc^M DNA

SUBSTITUTE SHEET

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US90/07333

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) 6

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(5): C07H 19/12; C12Q 1/68; C12N 15/00
U.S. Cl.: 536/27; 435/6; 935/77, 78

II. FIELDS SEARCHED

Minimum Documentation Searched 7

Classification System	Classification Symbols
U.S. Cl.	536/27; 435/6; 935/77, 78

Documentation Searched other than Minimum Documentation
to the Extent that such Documents are Included in the Fields Searched 8

APS, STN, Gen Bank, EMBL

III. DOCUMENTS CONSIDERED TO BE RELEVANT 9

Category*	Citation of Document, 11 with indication, where appropriate, of the relevant passages 12	Relevant to Claim No. 13
A	Journal of General Virology, Volume 63, issued 1982, Yang et al, "Molecular Cloning of the Endogenous Rat C-type Helper Virus DNA Sequence: Structural Organization and Functional Analysis of Same Restricted DNA fragments" pages 25-36. Some	1-23

* Special categories of cited documents: 10

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
28 January 1991	20 FEB 1991

International Searching Authority	Signature of Authorized Officer
ISA/US	Mindy B. Fleisher

Form PCT/ISA/210 (second sheet) (Rev.11-87)